Linear Algebra notes

Math 20F

August 2014

Contents

1 Linear systems, matrices, and vectors

Consider a linear system

$$\begin{cases} 2x_1 + x_2 + 5x_3 = -1 \\ x_1 + 6x_3 = 2 \\ -6x_1 + 2x_2 + 4x_3 = 3 \end{cases}$$

This can be written as:

1. an augmented matrix
$$\begin{bmatrix} 2 & 1 & 5 & | & -1 \\ 1 & 0 & 6 & | & 2 \\ -6 & 2 & 4 & | & 3 \end{bmatrix}$$

2. a matrix equation
$$\begin{bmatrix} 2 & 1 & 5 \\ 1 & 0 & 6 \\ -6 & 2 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \\ 3 \end{bmatrix}$$

3. a vector equation
$$x_1 \begin{bmatrix} 2 \\ 1 \\ -6 \end{bmatrix} + x_2 \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} + x_3 \begin{bmatrix} 5 \\ 6 \\ 4 \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \\ 3 \end{bmatrix}$$

all of which are equivalent representations.

Matrices. An $m \times n$ matrix has m rows and n columns.

Transpose of a matrix. The transpose of an $m \times n$ matrix A, A^T, is the $n \rtimes m$ matrix formed by turning the columns of A into the corresponding rows of A^T.

Note: $(AB)^T = B^T A^T$, **not** $A^T B^T$

Vectors. A column vector is an $n \times 1$ matrix, and a row vector is a $1 \times n$ matrix.

1.1 Row echelon and reduced row echelon form

A matrix can be put into row echelon form (ref) through basic row operations. Because of the order of the operations, ref need not be unique. The reduced row echelon form (rref), however, is unique.

Basic row operations

Replacement replace one row, R_i, with the sum of itself and the multiple of another, R_j.
$R_i \to R_i + cR_j$

Interchange Swap two rows. $R_i \leftrightarrow R_j$

Scaling multiply all entries of a wor by a nonzero number. $R_i \to cR_i$ $(c \neq 0)$

Pivots.

Basic and free variables. A variable corresponding to a pivot column is said to be basic, and a variable attached to a non-pivot column is free (since its value depends on that of basic variables).

ref: Gaussian elimination

rref: Gauss-Jordan elimination

Existence and uniqueness of solutions. A system can either be

1. **Consistent**, in which case it can either have
   - a unique solution (No free variables in the system)
   - an infinite number of solutions (free variables present in system).
2. **Inconsistent**. No solution exists, i.e.
   - a row in the rref of the augmented matrix is of the form $\begin{bmatrix} 0 & \ldots & 0 & c \end{bmatrix}$, where $c \neq 0$
   - the rightmost column of the augmented matrix is a pivot column.

2 Linear combinations

Definition 2.1 (Linear combination)**.** Given vectors $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_p}$ in $\mathbb{R}^n$ and corresponding scalars $c_1, c_2, \ldots, c_p$, the vector $\mathbf{y} = c_1\mathbf{v_1} + c_2\mathbf{v_2} + \ldots + c_p\mathbf{v_p}$ is a linear combination of $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_p}$ with weights $c_1, c_2, \ldots, c_p$.

Definition 2.2 (Subset of $\mathbb{R}^n$ spanned by $\mathbf{v_1} \ldots \mathbf{v_p}$)**.** The subset spanned (or generated) by $\mathbf{v_1} \ldots \mathbf{v_p}$, Span$\{\mathbf{v_1} \ldots \mathbf{v_p}\}$, is the set of all linear combinations of $\mathbf{v_1} \ldots \mathbf{v_p}$ given by any group of scalars $c_1 \ldots c_p$.

If $\mathbf{b}$ is spanned by Span$\{\mathbf{v_1} \ldots \mathbf{v_p}\}$, then $\mathbf{b} \in$ Span$\{\mathbf{v_1} \ldots \mathbf{v_p}\}$, which is the same as asking if $x_1\mathbf{v_1} + x_2\mathbf{v_2} + \ldots + x_p\mathbf{v_p} = \mathbf{b}$ has a solution.

The zero vector must be in any span.

2.1 Matrix equations

Any equation of the form $A\mathrm{x} = \mathbf{b}$ is a matrix equation, where A is the coefficient matrix. It's a convenient representation for linear combinations.

Matrix equation theorem
(assume A is an $m \times n$ matrix; all properties are equivalent)

1. For each $\mathbf{b}$ in $\mathbb{R}^m$, $A\mathbf{x} = \mathbf{b}$ has a solution, that is,
2. Each $\mathbf{b}$ in $\mathbb{R}^m$ is a linear combination of the columns of A.
3. The columns of A span $\mathbb{R}^m$.
4. A has a pivot position in every row.

2.2 Homogeneous linear systems

A system is homogeneous if it can be written as $A\mathbf{x} = \mathbf{0}$, where A is the coefficient matrix.

There must always be at least one solution, the **trivial solution**, $\mathbf{x} = \mathbf{0}$. If a nonzero vector $\mathbf{x}$ also satisfies the equation, then the system also has a **nontrivial solution**.

2.3 Non-homogeneous linear systems

2.4 Linear dependence and independence

Definition 2.3. A set of vectors $S = \{\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_p}\}$ is linearly independent if the equation $x_1\mathbf{v_1} + x_2\mathbf{v_2} + \ldots + x_p\mathbf{v_p} = \mathbf{0}$ has only the trivial solution; and linearly dependent otherwise.

Test for lin. independence by forming an augmented matrix $[\mathbf{v_1}\ \mathbf{v_2}\ \ldots\ \mathbf{v_p} \mid \mathbf{0}]$ (or just a coefficient matrix), row reducing, and picking out free variables.

3 Linear transformations

Definition 3.1 (Transformation)**.** A transformation $T : \mathbb{R}^n \mapsto \mathbb{R}^m$ maps from each vector $\mathbf{x} \in \mathbb{R}^n$ to a vector $T(\mathbf{x}) \in \mathbb{R}^m$. $\mathbb{R}^n$ is the **domain**, $\mathbb{R}^m$ is the **codomain**, $T(\mathbf{x})$ is the **image** of $\mathbf{x}$, and the set of all images is the **range**.

Definition 3.2 (Linear transformation)**.** A transformation is *linear* if it satisfies the following:

1. $T(\mathbf{v} + \mathbf{u}) = T(\mathbf{v}) + T(\mathbf{u})$
2. $T(c\mathbf{u}) = cT(\mathbf{u})$

   $\Rightarrow T(\mathbf{0}) = \mathbf{0}$

$\Rightarrow T(c\mathbf{v} + d\mathbf{u}) = cT(\mathbf{v}) + dT(\mathbf{u})$

When we multiply a matrix A with a vector $\mathbf{x}$, A transforms $\mathbf{x}$ into the resulting vector $\mathbf{b}$. The solution to $A\mathbf{x} = \mathbf{b}$ is therefore the set of all $\mathbf{x}$ in $\mathbb{R}^n$ that are transformed to $\mathbf{b}$ in $\mathbb{R}^m$. This *matrix transformation* is denoted $\mathbf{x} \mapsto A\mathbf{x}$, or $T(\mathbf{x}) = A\mathbf{x}$.

Standard matrix. The standard matrix A of a linear transformation $T : \mathbb{R}^n \mapsto \mathbb{R}^m$ s.t. $T(\mathbf{x}) = A\mathbf{x}$ is an $m \times n$ matrix formed from the transformation of each vector in the standard basis for $\mathbb{R}^n$ (i.e. the columns of the $n \times n$ identity matrix). $A = [T(\mathbf{e}_1) \ \ldots \ T(\mathbf{e}_n)]$

To recover the standard matrix given a transformation, say

$$T : \mathbb{R}^3 \mapsto \mathbb{R}^2 \Rightarrow T(x_1, x_2, x_3) = (x_1 - 5x_2 + 4x_3, x_2 - 6x_3)$$

the rows of the matrix are comprised of the coefficients of each entry in the resulting vector, so

$$A = \begin{bmatrix} 1 & -5 & 4 \\ 0 & 1 & -6 \end{bmatrix}$$

Injections and surjections

one-to-one A transformation is injective if each $\mathbf{b} \in \mathbb{R}^m$ is the image of at *most* one $\mathbf{x} \in \mathbb{R}^n$.
Test by seeing if the columns of A are linearly independent, i.e. if $T(\mathbf{x}) = \mathbf{0}$ has only the trivial solution.

onto A transformation is surjective if each $\mathbf{b} \in \mathbb{R}^m$ is the image of at *least* one $\mathbf{x} \in \mathbb{R}^n$.
Test by seeing if the columns of A span $\mathbb{R}^m$, i.e. if there are pivot positions in every row of A.

4 Inverse of a matrix

Invertible matrix theorem
Let A be an $n \times n$ matrix. The following statements are all equivalent.

1. A is invertible
2. If A is invertible, then A^T is also invertible.
3. A is row equivalent to I_n
4. A is column equivalent to I_n
5. A has n pivot positions
6. $A\mathbf{x} = 0$ has only the trivial solution
7. Columns of A are linearly independent
8. Columns of A span $\mathbb{R}^n$
9. $A\mathbf{x} = \mathbf{b}$ has at least one soluton for each $\mathbf{b} \in \mathbb{R}^n$
10. $x \mapsto Ax$ is one-to-one
11. $x \mapsto Ax$ maps from $\mathbb{R}^n$ to $\mathbb{R}^n$.
12. There exists matrices C, D such that $CA = AD = I$
13. The columns of A form a basis for $\mathbb{R}^n$
14. Col$(A) = \mathbb{R}^n$
15. dim(Col(A)) = Rank$(A) = n$
16. Nul$(A) = \{\mathbf{0}\}$
17. dim(Nul(A)) $= 0$

If we have a matrix equation $A\mathbf{x} = \mathbf{b}$, then $\mathbf{x} = A^{-1}\mathbf{b}$ has a unique solution if A is invertible. Otherwise, if A is singular (non-invertible, det$(A) = 0$), there are no real solutions.

Note: $(AB)^{-1} = B^{-1}A^{-1}$

Finding the inverse.
Let A be an $n \times n$ matrix. Notice that if A is invertible, A is row equivalent to I_n. Therefore, the same row operations that will reduce A to rref, when applied to I_n, will give A^{-1}.

To find A^{-1}, simply put A and I_n into an augmented matrix and perform Gauss-Jordan elimination on the A submatrix. When the LHS is reduced to I, the RHS will equal A^{-1}.

If A cannot be row reduced to I, then A is singular.

$$\begin{bmatrix} A & | & I_n \end{bmatrix} \xrightarrow{\text{rref on A}} \begin{bmatrix} I_n & | & A^{-1} \end{bmatrix}$$

Or, use the more tedious formula $A^{-1} = \frac{1}{\det(A)} C^T$ (transpose of the cofactor matrix).

Special case: if A is a 2×2 matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, then $A^{-1} = \frac{1}{\det(A)} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$

5 Determinants

Properties of the determinant
(Only defined for $n \times n$ matrices.)

1. $\det(I) = 1$
2. Swapping two rows reverses the sign of the determinant.
3. The determinant is a linear function for each row separately, given that the other rows remain the same. E.g.
4. If A has two equal rows, $\det(A) = 0$
5. Row replacement does not change the determinant.
6. If A has a zero row, $\det(A) = 0$
7. If A is a triangular matrix, then $\det(A)$ is the product of the diagonal entries.
8. $\det(A) = 0$ if A is singular
9. $\det(AB) = \det(A)\det(B)$, which implies that $\det(A^{-1}) = 1/\det(A)$
10. $\det(A) = \det(A^T)$ (Implies that the above operations can also be performed on the columns to the same effect.)

5.1 Naïve method: cofactor expansion

Minor: The minor of a matrix at row i and column j, M_{ij}, is the determinant of the matrix formed by removing the i^{th} row and j^{th} column.

Cofactor: The cofactors of a matrix are found by $C_{ij} = (-1)^{i+j} M_{ij}$. A cofactor matrix C is the matrix formed by the cofactors at each position.

Adjoint: The adjoint of a matrix is the transpose of the cofactor matrix $\text{adj}(A) = C^T$.

Cofactor expansion: Can be done down any column or row:

$$\det(A) = \begin{cases} \sum_{k=1}^{n} a_{ik}C_{ik}, \ i^{th} \text{ row} \\ \sum_{k=1}^{n} a_{kj}C_{kj}, \ j^{th} \text{ column} \end{cases}$$

Choose a row or column (preferably with the most zeroes), and sum the minors of each element down that row/column.

Cramer's rule

To solve the nonhomogeneous equation $A\mathbf{x} = \mathbf{b}$ for $\mathbf{x}$. If A is an invertible $n \times n$ matrix, and $\mathbf{b} \in \mathbb{R}^n$, then the entries of $\mathbf{x}$ are given by

$$x_i = \frac{\det A_i(\mathbf{b})}{\det A} \quad (1 \leq i \leq n)$$

where $\det A_i(\mathbf{b})$ is found by replacing the i^{th} column of A with $\mathbf{b}$.

Shapes in Euclidean space

Let $\mathbf{a}, \mathbf{b}$ be vectors (sides of the parallelogram) in $\mathbb{R}^2$. Then the area of the parallelogram enclosed is $\det([\mathbf{a}\ \mathbf{b}])$. Similarly, if $\mathbf{a}, \mathbf{b}, \mathbf{c}$ are in $\mathbb{R}^3$, then the area of the parallelopiped enclosed is $\det([\mathbf{a}\ \mathbf{b}\ \mathbf{c}])$. (Works for both column and row vectors, just form an $n \times n$ matrix.)

6 Vector spaces

Definition 6.1 (Vector space)**.** A vector space is a set of vectors (more generally, objects) V with vector addition and scalar multiplication, such that the operations satisfy the all of the following properties ($\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ and $c, d \in \mathbb{R}$):

Addition in $\mathbb{R}^n$

1. Closure property. $(\mathbf{u} + \mathbf{v}) \in V$
2. Commutativity. $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$
3. Associativity. $(\mathbf{u}+\mathbf{v})+\mathbf{w} = \mathbf{u}+(\mathbf{v}+\mathbf{w})$
4. Additive identity. $\mathbf{0} + \mathbf{u} = \mathbf{u} + \mathbf{0} = \mathbf{u}$
5. Aditive inverse. $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$

Scalar multiplication

6. Closure property. $(c\mathbf{u}) \in V$
7. Distributivity I. $c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v}$
8. Distributivity II. $(c + d)\mathbf{u} = c\mathbf{u} + d\mathbf{u}$
9. Associativity. $(cd)\mathbf{u} = c(d\mathbf{u})$
10. Multiplicative identity. $1\mathbf{u} = \mathbf{u}$

Examples of vector spaces.

Yes:

- $\mathbf{M_2}$, the set of all 2×2 matrices

- $\mathbb{P}_2$, the set of all polynomials of degree ≤ 2
- All real valued continuous functions

  No:

- $\mathbb{Z}$, the set of all integers
- $\mathbb{R}^2$, with scalar multiplication redefined as $c\left[\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right] = \left[\begin{smallmatrix} cx_1 \\ 0 \end{smallmatrix}\right]$
- $\mathbb{P}_2$, the subset of all polynomials of degree $= 2$

6.1 Subspace of a vector space

Definition 6.2 (Subspace of a vector space)**.** For the vector spaces V, W, if $W \subseteq V$, W is a subset of V if the following hold: (a) $\mathbf{0}_n \in W$, (b) W is closed under addition and scalar multiplication.

Theorem 6.1 (Subspace spanned by a set)**.** *If* $\mathbf{b_1}, \dots, \mathbf{b_n}$ *are in a vector space* V*, then Span*$\{\mathbf{b_1}, \dots, \mathbf{b_n}\}$ *is a subspace of* V*.*

6.2 Null, column, and row space

	Nul(A)	Col(A)
Defn	$\{\mathbf{x} \in \mathbb{R}^n : \ A\mathbf{x} = \mathbf{0}\}$	Span(cols. of A) $\{\mathbf{b} \in \mathbb{R}^m : \ A\mathbf{x} = \mathbf{b} \ \ \exists \mathbf{x} \in \mathbb{R}^m\}$
Subspace of	$\mathbb{R}^n$	$\mathbb{R}^m$
Check if $\mathbf{v}$ is in	Check if $A\mathbf{v} = \mathbf{0}$	Solve $A\mathbf{x} = \mathbf{v}$ for $\mathbf{x}$ $\Rightarrow$ see if $[A \ \mathbf{v}]$ is consistent
Find spanning set	Solve $A\mathbf{x} = \mathbf{0}$ $\Rightarrow$ write $\mathbf{x}$ as a vector equation $\mathbf{x} = \mathbf{v_1}x_2 + \mathbf{v_2}x_2 + \ldots + \mathbf{v_n}x_n$ If no free variables, then Nul$(A) = \{\mathbf{0}_n\}$. Else, Nul$(A) =$ Span$\{\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_n}\}$	Span$\{a_1, a_2, \ldots, a_n\}$

The null space is also known as the kernel when talking about linear ternsformations. The kernel of a transformation T is the set of all $\mathbf{u} \in V$ such that $T(\mathbf{u}) = \mathbf{0}$.

Row space

Definition 6.3 (Row space)**.** The vector space spanned by the rows of A, Row(A) = Span$\{r_a, \ldots, r_n\}$. If A is an $m \times n$ matrix, then Row$(A) \subseteq \mathbb{R}^n$.

Elementary row operations preserve the null space and row space, so if $A \xrightarrow{\text{ref}} U \xrightarrow{\text{rref}} R$, then Row$(A)$ = Row(U) = Row(R).

6.3 Bases and dimensions

A basis is a minimal spanning set (spanning the entire space with the least possible number of vectors) for a vector space.

Definition 6.4 (Basis)**.** A set of vectors $S = \{\mathbf{v_1}, \ldots, \mathbf{v_p}\}$ in a vector space V is a basis for V if: (a) S is linearly independent (so for $\mathbb{R}^n$, the basis is formed by n vectors), (b) $V = \text{Span}(S)$.

Theorem 6.2 (Basis Theorem)**.** *Let H be a p-dimensional subspace for $\mathbb{R}^n$. Any linearly independent set of **exactly** p elements is a basis for H. Also, any set of p elements that spans H is a basis for H.*

Standard and non-standard bases.
$S = \{[1,0,0]^T, [0,1,0]^T, [0,0,1]^T\}$ and $T = \{[1,2,3]^T, [0,1,2]^T, [2,0,1]^T\}$ both form a basis for $\mathbb{R}^3$. S is a standard matrix, while T is a non-standard matrix. In general, the standard basis (for a Euclidian space) is formed by the set of unit vectors pointing in the direction of the axes in Cartesian coordinates. Standard bases for other vector spaces are found using similar logic (e.g. for $\mathbb{P}_2$, it's $\{1, t, t^2\}$).

To test if a set of n vectors is in $\mathbb{R}^n$

1. Put the vectors in a matrix A.
2. If A is invertible (therefore lin. indep), then the vectors form a basis.

   Similarly, if $A\text{x} = \mathbf{0}$ has only the trivial solution, the vectors form a basis.

Theorem 6.3. *If a vector space V contains a basis $\mathcal{B} = \{\mathbf{b_1}, \ldots, \mathbf{b_n}\}$, then (a) all sets of more than n vectors are linearly dependent (b) every other basis must bave* exactly *n vectors.*

Theorem 6.4 (Spanning set theorem)**.** *Let $S = \{\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_p}\}$ be vectors in V, and $H =$ Span(S). If $\mathbf{v_k} \in S$ is a linear combination of other vectors in S, then $S' = \{S - \{\mathbf{v_k}\}\}$ still spans H.*

Bases for null, column, and row spaces

1. Basis for Nul(A)
   - Let A be an $m \times n$ matrix, and let R be the rref form of A.
   - To find the basis for Nul(A), solve

$$\begin{aligned} A\mathbf{x} &= \mathbf{0} \\ \Rightarrow R\mathbf{x} &= \mathbf{0} \end{aligned}$$

   - Read off R to find the general solution as a vector equation in terms of free variables $\mathbf{x} = \mathbf{v_1}x_a + \mathbf{v_2}x_b + \ldots + \mathbf{v_n}x_n$, where $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_n}$ are the "special solutions".

  - The set of special solutions is the basis for $\text{Nul}(A)$, where $\text{Nul}(A) = \text{Nul}(R)$.
  - If R has no free variables, $\text{Nul}(A) = \{\mathbf{0}\}$

2. Basis for $\text{Col}(A)$
  - *Note: Col$(A) \neq$ Col(R)*
  - $\text{Col}(R) = \text{Span}\{\text{columns of } R\}$.
  - Take the columns of R, remove linearly depenedent vectors by the spanning set theorem, and the remaining is still a basis for R. But notice that the lin. dep. columns are the non-pivot columns.
  - Thus, the pivot cols of R form a basis for $\text{Col}(R)$.
  - But the pivot column positions in A are the same as R, so by inspecting R, we can find the pivot columns of A, and thereby $\text{Col}(A)$.

3. Basis for $\text{Row}(A)$
  - Find a set of linearly independent row vectors in $R \Rightarrow$ remove zero rows of $\text{Row}(R)$.
  - The basis of $\text{Row}(A) = \{\text{nonzero rows in } R\}$. (Or the corresponding rows in A.)

Dimension

Definition 6.5 (Dimension of a vector space)**.** The dimension of a vector space is $|\mathcal{B}|$, with the special case that $\dim(\{\mathbf{0}\}) = 0$.

(A is an $m \times n$ matrix)

- Nullity of A: dim(Nul(A)) = number of non-pivot columns
- Rank(A) = dim(Col(A)) = dim(Row(A)) = number of pivot columns
- Nullity(A) + Rank(A) = n
- Row(A) = Col(A^T); Col(A) = Row(A^T)

6.4 Coordinates and bases

Geometrically, the basis can be used as a coordinate system, as its linear combinations can be used to produce every other vector in a given vector space V.

Suppose we have a vector $\mathbf{x} \in V$ that is expressed in some basis (assume it's the standard basis). We can write the same vector in terms of another basis $\mathcal{B}$ by

$$\mathbf{x} = P_{\mathcal{B}}[\mathbf{x}]_{\mathcal{B}}$$
$$[\mathbf{x}]_{\mathcal{B}} = {P_{\mathcal{B}}}^{-1}\mathbf{x}$$

where $P_{\mathcal{B}}$ is the $\mathcal{B}$-coordinate matrix formed by the vectors in $\mathcal{B}$, and $[\mathbf{x}]_{\mathcal{B}}$ is the vector $\mathbf{x}$ relative to $\mathcal{B}$.

7 Eigenvectors and eigenvalues

$$A\mathbf{x} = \lambda\mathbf{x}$$

The eigenvector is a nonzero vector $\mathbf{x}$ and the eigenvalue is a corresponding scalar λ that satisfies the above equation for some invertible matrix A. (All matrices A mentioned in this section are $n \times n$.)

Note: Row operations change the eigenvalues.

Characteristic equation

$$\overbrace{\underbrace{\det(A - \lambda I)}_{\text{characteristic polynomial}} = 0}^{\text{characteristic equation}}$$

λ is an eigenvalue of A if and only if the characteristic equation is satisfied.

1. To check if $\mathbf{y}$ is an eigenvector of A, solve $A\mathbf{y}$ and check if the result is a scalar multiple of $\mathbf{y}$.
2. To check if $\lambda = c$ is an eigenvalue of A, substitute c into the characteristic equation and check if the determinant is 0.
   2.1. To find the corresponding eigenvector, solve for $\text{Nul}(A - \lambda I)$
   2.2. The set of eigenvectors corresponding to an eigenvalue is an **eigenspace**.
3. To find all eigenvalues of A, solve the characteristic polynomial for λ. (Write $A - \lambda I$ as a matrix with the unknown λ, expand the determinant, factorize.)
   - In the case of repeated roots, the eigenvalue R has multiplicity n.

Theorem 7.1. *Eigenvectors corresponding to distinct eigenvalues are linearly independent.*

Powers of matrices and eigenvalues

$$\begin{aligned} A\mathbf{x} &= \lambda\mathbf{x} \\ \Rightarrow AA\mathbf{x} &= A\lambda\mathbf{x} \\ &= \lambda(A\mathbf{x}) \\ &= \lambda(\lambda\mathbf{x}) \\ A^2\mathbf{x} &= \lambda^2\mathbf{x} \end{aligned}$$

and in general, $A^k\mathbf{x} = \lambda^k\mathbf{x}$. (The eigenvalue corresponding to A^k is λ^k, the eigenvector remains the same.)

7.1 Diagonalization

We want to write A in a way that allows us to compute A^k easily. Recall that a diagonal matrix raised to a power k is just the same matrix with all the diagonal entries raised to k. If we write A in the diagonalized form $A = PDP^{-1}$ where D is a diagonal matrix, then

$$\begin{aligned} A^k &= (PDP^{-1})^k \\ &= \underbrace{(PDP^{-1})(PDP^{-1})\dots(PDP^{-1})}_{k \text{ times}} \\ &= P\underbrace{DD\dots D}_{k \text{ times}}P^{-1} \\ &= PD^kP^{-1} \end{aligned}$$

We form P and D from the eigenvectors and eigenvalues. Because P must be invertible, it follows from theorem 7.1 that the $n \times n$ matrix A must have n distinct eigenvalues for it to be diagonalizable.

$$P = [\mathbf{v_1}\ \mathbf{v_2} \dots \mathbf{v_n}] \qquad \text{where } \mathbf{v_1}, \mathbf{v_2}, \dots, \mathbf{v_n} \text{ are the eigenvectors}$$

$$D = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} \qquad \text{(the rest of the matrix are zeroes)}$$

Definition 7.1 (Similarity)**.** A and B are similar if $A = PBP^{-1}$ for some invertible matrix P. $\Rightarrow$ A and B have the same eigenvalues.

Orthogonal diagonalization and symmetric matrices

Properties of symmetric matrices

- A symmetric matrix is a square matrix A such that $A^T = A$.
- If A is symmetric, then all eigenvalues of A are real numbers.
- Any two eigenvectors from different eigenspaces are orthogonal.

Spectral Theorem for Symmetric matrices
An $n \times n$ symmetric matrix A has the following properties

1. A has n real eigenvalues, counting multiplicity.
2. Eigenvectors corresponding to distinct eigenvalues are orthogonal, so all eigenspaces are mutually orthogonal.
3. The dimension of the eigenspace for each eigenvalue λ is the multiplicity of λ.
4. A is orthogonally diagonalizable.

Orthogonal diagonalization. If A is symmetric, then we can take a set of eigenvectors that are orthogonal and form an orthogonal matrix U. Then, $A = UDU^{-1} = UDU^T$

Theorem 7.2. *A square matrix A is orthogonally diagonalizable if and only if A is a symmetric matrix.*

8 Orthogonality, projections, least squares

Recap: Vector operations and geometry

1. $\mathbf{u}\cdot\mathbf{v} = \sum u_i v_i = \|\mathbf{u}\|\|\mathbf{v}\|\cos\theta$ and $\mathbf{u}\times\mathbf{v} = \begin{vmatrix} 1 & 1 & 1 \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix} = \|\mathbf{u}\|\|\mathbf{v}\|\sin\theta$ (cross product only defined if $\mathbf{u}, \mathbf{v} \in \mathbb{R}^3$).
2. Vectors are perpendicular if $\mathbf{u}\cdot\mathbf{v} = 0$ and parallel if $\mathbf{u}\times\mathbf{v} = \vec{0}$.
3. The magnitude or norm of a vector, $\|\mathbf{u}\| = \sqrt{\mathbf{u}\cdot\mathbf{u}}$. Distance between two vectors $\mathrm{d}(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|$.
4. The unit vector, or normalized vector (of a vector $\mathbf{u}$) is $\hat{\mathbf{u}} = \frac{\mathbf{u}}{\|\mathbf{u}\|}$.

8.1 Orthogonal sets

Definition 8.1 (Orthogonal set)**.** A set $S = \{\mathbf{v_1}, \ldots, \mathbf{v_p}\}$ is an orthogonal set if every vector in it is orthogonal to every other vector ($\mathbf{v_i}\cdot\mathbf{v_j} = 0\ \forall i \neq j$).

Definition 8.2 (Orthogonal basis)**.** If S is an orthogonal set, then the vectors in S are linearly independent (meaning it can form a basis). S forms an orthogonal basis for V if it is also a basis for a subspace V.

Definition 8.3 (Orthogonal complement)**.** If $\mathbf{x}$ is orthogonal to all vectors in V, then $\mathbf{x}$ is orthogonal to V. The set of all such $\mathbf{x}$ is the orthogonal complement, $V^\perp$, of V.

In particular, $\text{Row}(A)^\perp = \text{Nul}(A)$ (in $\mathbb{R}^n$), and $\text{Col}(A)^\perp = \text{Nul}(A^T)$ (in $\mathbb{R}^m$).

Example: $S = \{\mathbf{v_1}, \ldots, \mathbf{v_p}\}$. To find $S^\perp$, write a matrix $A = [\mathbf{v_1}\ \ldots\ \mathbf{v_p}]$. Observe that $S = \text{Col}(A)$, so $S^\perp = \text{Nul}(A^T)$.

Theorem 8.1 (Orthogonality $\Rightarrow$ linear independence)**.** *If S is a set of orthogonal vectors, then S is linearly independent.*

Definition 8.4 (Orthonormal set, orthogonal matrix)**.** If S is an orthogonal set, and if all vectors in S are normalized such that $\|\mathbf{u_i}\| = 1$ for all i, then S is an orthonormal set. The matrix formed from the elements of the orthonormal set, $U = [\mathbf{u_1}\ \ldots\ \mathbf{u_n}]$ (where $\mathbf{u_i} \in \mathbb{R}^m$) is an *orthogonal* matrix.

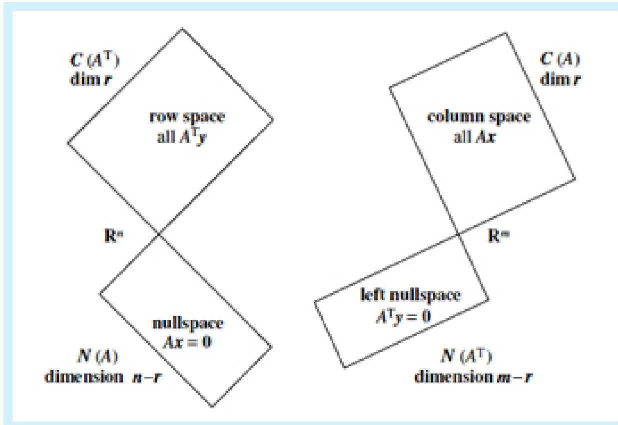

Figure 1: Orthogonality of the fundamental subspaces of $\mathbb{R}$

Properties of U: If U is an $m \times n$ orthogonal matrix, then

1. $U^TU = I_n$ and $UU^T = I_m$
2. $\|U\mathbf{x}\| = \|\mathbf{x}\| \; \forall \mathbf{x} \in \mathbb{R}^n$
3. $$\begin{aligned}(U\mathbf{x}) \cdot (U\mathbf{y}) &= (U \cdot \mathbf{y})^T(U \cdot \mathbf{x}) \\ &= \mathbf{y}^TU^TU\mathbf{x} \\ &= \mathbf{xy}\end{aligned}$$
4. $(U\mathbf{x}) \cdot (U\mathbf{y}) = 0$ iff $\mathbf{x} \cdot \mathbf{y} = 0$

   (the linear transform $\mathbf{x} \mapsto U\mathbf{x}$ preserves length and orthogonality of $\mathbf{x}$.)
5. For a **square** orthogonal matrix, $U^{-1} = U^T$

8.2 Orthogonal decomposition

Theorem 8.2 (Weights from an orthogonal basis)**.** *Let S be an orthogonal basis of a subspace $W \in \mathbb{R}^n$ that gives $\mathbf{y} \in W$. The weights of the linear combination $\mathbf{y} = c_1\mathbf{v_1} + c_2\mathbf{v_2} + \ldots + c_p\mathbf{v}_p$, for any $y \in W$, is given by*

$$c_i = \frac{\mathbf{y} \cdot \mathbf{v}_i}{\mathbf{v}_i \cdot \mathbf{v}_i}$$

Proof. Take the linear combination $\mathbf{y} = c_1\mathbf{v_1} + c_2\mathbf{v_2} + \ldots + c_p\mathbf{v}_p$. For every $1 \leq i \leq p$, compute $\mathbf{y} \cdot \mathbf{v}_i$. Because S is an orthogonal basis, all vectors drawn from the basis are orthogonal to every other (and their dot product is 0). Thus, $\mathbf{y} \cdot \mathbf{v}_i = c_i(\mathbf{v}_i \cdot \mathbf{v}_i)$. Rearranging, we have $c_i = \frac{\mathbf{y} \cdot \mathbf{v}_i}{\mathbf{v}_i \cdot \mathbf{v}_i}$, for every i. □

Orthogonal projection. Given two vectors, $\mathbf{a}$ and $\mathbf{b}$ in $\mathbb{R}^n$, we wish to decompose one of them such that $\mathbf{b} = \hat{\mathbf{b}} + \mathbf{e}$, where $\hat{\mathbf{b}}$ is a multiple of $\mathbf{a}$ and $\mathbf{e}$ is orthogonal to $\mathbf{a}$. From theorem 8.2, Span$\{\mathbf{a}\}$ forms an orthogonal basis for the orthogonal projection of $\mathbf{b}$ onto $\mathbf{a}$, so we can find the scaling factor with the equation above.

$$\hat{\mathbf{b}} = \text{proj}_{\text{Span}\{\mathbf{a}\}}\mathbf{b} = \frac{\mathbf{b}\cdot\mathbf{a}}{\mathbf{a}\cdot\mathbf{a}}\mathbf{a}$$

The component of $\mathbf{b}$ orthogonal to $\mathbf{a}$, $\mathbf{e}$, can be found by simple algebra: $\mathbf{e} = \mathbf{b} - \hat{\mathbf{b}}$.

Theorem 8.3 (Orthogonal decomposition)**.** *Extending the above example, given a subspace $W \in \mathbb{R}^n$ with an orthogonal basis $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$, each $\mathbf{y} \in \mathbb{R}^n$ can then be written as a sum $\mathbf{y} = \hat{\mathbf{y}} + \mathbf{e}$, where $\hat{\mathbf{b}} \in W$ and the error $\mathbf{e} \in W^{\perp}$. The projection $\hat{\mathbf{y}}$ is given as $proj_W\mathbf{y} = c_1\mathbf{v}_1 + \ldots + c_p\mathbf{v}_p$ where $c_i = \frac{\mathbf{y}\cdot\mathbf{v}_i}{\mathbf{v}_i\cdot\mathbf{v}_i}$ for each $1 \leq i \leq p$, and the error is again $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}$.*

Distance from point to line. This yields a nice alternative way of computing the distance from a point to a line (cf. vector geometry with parametric forms...) — the distance is equivalent to the magnitude of the error vector. Treat the point as a line from the origin, find the projection of the point to the vector, then the magnitude of the error vector.

The Gram-Schmidt process: finding an orthogonal basis
Given a vectorspace V with a basis $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$, an orthogonal basis $\mathcal{B}' = \{\mathbf{w}_1, \ldots, \mathbf{w}_p\}$ can be found by

$$\begin{aligned}
\mathbf{w}_1 &= \mathbf{v}_1 \\
\mathbf{w}_2 &= \mathbf{v}_2 - \frac{\mathbf{v}_2\cdot\mathbf{w}_1}{\mathbf{w}_1\cdot\mathbf{w}_1}\mathbf{w}_1 \\
\mathbf{w}_3 &= \mathbf{v}_3 - \frac{\mathbf{v}_3\cdot\mathbf{w}_1}{\mathbf{w}_1\cdot\mathbf{w}_1}\mathbf{w}_1 - \frac{\mathbf{v}_3\cdot\mathbf{w}_2}{\mathbf{w}_2\cdot\mathbf{w}_2}\mathbf{w}_2 \\
&\vdots \\
\mathbf{w}_p &= \mathbf{v}_p - \left[\sum_{i=1}^{p-1}\frac{\mathbf{v}_p\cdot\mathbf{w}_i}{\mathbf{w}_i\cdot\mathbf{w}_i}\mathbf{w}_i\right]
\end{aligned}$$

If we normalize each vector in $\mathcal{B}'$, we will have an orthonormal basis $\mathcal{B}''$

8.3 Least-squares problems

Say we want to find a linear regression through a dataset — this is equivalent